Novel Insights into the Landscape of Crossover and Noncrossover Events in Rhesus Macaques (*Macaca mulatta*)

Cyril J. Versoza[1,2,†], Sarah Weiss[1,†], Ravneet Johal[1], Bruno La Rosa[1], Jeffrey D. Jensen[1,2], and Susanne P. Pfeifer [1,2,*]

[1]School of Life Sciences, Arizona State University, Tempe, AZ, USA

[2]Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: susanne@spfeiferlab.org.

Accepted: November 28, 2023

Abstract

Meiotic recombination landscapes differ greatly between distantly and closely related taxa, populations, individuals, sexes, and even within genomes; however, the factors driving this variation are yet to be well elucidated. Here, we directly estimate contemporary crossover rates and, for the first time, noncrossover rates in rhesus macaques (*Macaca mulatta*) from four three-generation pedigrees comprising 32 individuals. We further compare these results with historical, demography-aware, linkage disequilibrium–based recombination rate estimates. From paternal meioses in the pedigrees, 165 crossover events with a median resolution of 22.3 kb were observed, corresponding to a male autosomal map length of 2,357 cM—approximately 15% longer than an existing linkage map based on human microsatellite loci. In addition, 85 noncrossover events with a mean tract length of 155 bp were identified—similar to the tract lengths observed in the only other two primates in which noncrossovers have been studied to date, humans and baboons. Consistent with observations in other placental mammals with PRDM9-directed recombination, crossover (and to a lesser extent noncrossover) events in rhesus macaques clustered in intergenic regions and toward the chromosomal ends in males—a pattern in broad agreement with the historical, sex-averaged recombination rate estimates—and evidence of GC-biased gene conversion was observed at noncrossover sites.

Key words: meiosis, recombination, gene conversion, primate genomics.

Significance

Meiotic recombination is initiated by the programmed formation of double-strand breaks, which are resolved either with (crossover) or without (noncrossover) an exchange of flanking markers. Previous work across the tree of life has demonstrated that crossover landscapes are highly variable at every scale examined—however, detailed knowledge of noncrossover landscapes remains elusive in many species. Based on genomic data from four three-generation pedigrees, we here characterize crossover and, uniquely, noncrossover events in rhesus macaque (*Macaca mulatta*)—the most important nonhuman primate model in biomedical research—and compare our findings with recent observations in humans and another catarrhine monkey, baboons.

Introduction

Meiotic recombination is a tightly controlled process crucial for successful gametogenesis in nearly all sexually reproducing organisms, as it is required for the proper pairing and segregation of homologous chromosomes (Cole et al. 2012; Baudat et al. 2013). In addition, it plays a major role in shaping genetic diversity in populations by providing novel combinations of alleles for natural selection to act upon. Specifically, by shuffling parental alleles, recombination can improve the efficacy of selection by bringing beneficial alleles at different loci onto a common genetic background or by breaking down the linkage between beneficial and deleterious alleles (Hill and Robertson 1966). Recombination also facilitates the purging of deleterious alleles from populations, thus reducing the mutational load (Muller 1964; Felsenstein 1974), and its rate modulates the effects of both genetic hitchhiking and background selection acting across the genome (Maynard Smith and Haigh 1974; Charlesworth et al. 1993; and see the review of Charlesworth and Jensen 2021).

Recombination is initiated in prophase I of meiosis by the programmed formation of double-strand breaks (DSBs) between homologous chromosomes catalyzed by the protein SPO11 (Keeney 2001), the repair of which can lead to either a reciprocal exchange of chromatid arms between homologs—termed "crossover" (CO)—or a unidirectional replacement of a small genomic region in one of the chromatids, while leaving the donor homolog unchanged—termed "noncrossover" (NCO) (as illustrated in fig. 1*a*). In most organisms, at least one CO per chromosome (or chromosome arm) is necessary for a balanced segregation of homologous chromosomes (Jones and Franklin 2006; Coop and Przeworski 2007)—a so-called "crossover assurance"—as the absence of a faithful CO can be highly deleterious, often resulting in chromosomal nondisjunction and aneuploidy in species without an alternative mechanism (Hassold and Hunt 2001).

CO rates tend to vary between chromosomes of different lengths; in humans, for example, CO rate estimates range from 0.96 centimorgan per megabase (cM/Mb) in the longest chromosome (chromosome 1) to 2.11 cM/Mb in the shortest autosome (chromosome 22), with an overall sex-averaged rate of 1.1 cM/Mb (Kong et al. 2002). CO rates also vary considerably across different genomic regions (Buard and de Massy 2007), with low rates observed in CO deserts such as centromeric regions (<1 cM/Mb) and high rates observed in CO jungles such as subtelomeric regions (up to 5–10 cM/Mb; Clark et al. 2010). The majority of CO events occur in narrow (1–2 kb) regions of the genome—often referred to as "hotspots" due to an increased CO rate relative to the background (Kauppi et al. 2004)—that are targeted by the protein PRDM9 in primates as well as in many other mammals (Baudat et al. 2010; Myers et al. 2010; Parvanov et al. 2010; and see the review of Baudat et al. 2013). This leads to a block-like pattern of haplotypes that consists of long tracts with high levels of linkage disequilibrium (LD) interspersed with short tracts of low LD in many species (Guryev et al. 2006). Hotspots are highly dynamic, and heterogeneity in both their genome-wide distribution and intensity contributes to the tremendous variation in the number and distribution of COs that have been observed between distantly and closely related taxa, among different populations of the same species, as well as between individuals and sexes (see the review of Stapley et al. 2017).

In addition to COs, NCO events play an important role in shaping population diversity (Przeworski and Wall 2001). In fact, cytological, sperm-typing, and population genetic studies have suggested that the majority of DSBs culminate in NCOs, with about 1-fold to 15-fold more NCOs than COs depending on the taxa and genomic region (Jeffreys and May 2004; Baudat and de Massy 2007; Cole et al. 2010; Comeron et al. 2012; Li et al. 2019). Yet, in contrast to COs, which involve the readily identifiable exchange of large (often multiple kb long) stretches of DNA, detailed knowledge of NCO landscapes remains, with a few exceptions in model organisms (such as the flowering plant *Arabidopsis thaliana;* Sun 2012), elusive in many species. This is in part due to the short tract lengths generated (in humans, e.g., the mean tract length is 55–290 bp; Jeffreys and May 2004; Odenthal-Hesse et al. 2014) that make them difficult to disentangle from sequencing and genotyping errors—although both long and complex tracts, involving converted and nonconverted polymorphisms longer than 1 kb and extending up to 100 kb, have also been observed in human and nonhuman primates (Williams et al. 2015; Halldorsson et al. 2016; Wall et al. 2022). Moreover, in order for a NCO event to be identifiable, a polymorphism must be present in the donor homolog; thus, in species with low rates of heterozygosity, NCO tracts are frequently undetectable, as donor and recipient strands are identical (Wall et al. 2022).

Genomic approaches to study recombination involve the use of genome-wide genotype or sequencing data obtained from either related or unrelated individuals from a population sample (see the review of Clark et al. 2010). Tracking changes in haplotype structure across generations (fig. 1*b*), the pedigree-based approach allows for the direct investigation of contemporary CO and NCO rates in males and females separately, although often at low resolution, given the relatively small number of meiotic exchanges studied. In contrast, population-based approaches rely on coalescent theory to indirectly infer historical recombination rates from patterns of LD observed over many ancestral generations. LD-based methods offer higher resolutions over longer evolutionary timescales than classical pedigree approaches—however, inferred rates are necessarily

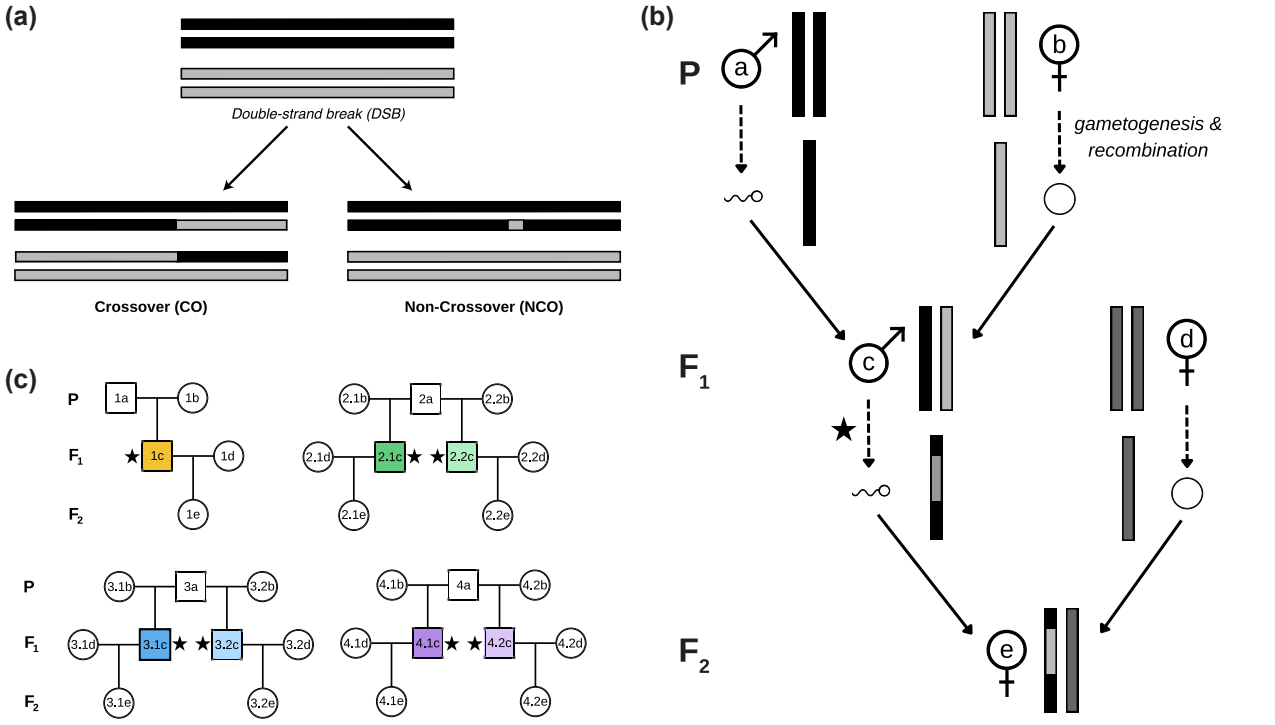

Fig. 1.—Meiotic recombination in the rhesus macaque pedigrees. (*a*) Recombination is initiated in prophase I of meiosis by the programmed formation of double-strand breaks between homologous chromosomes (shown in black and light gray), the repair of which can lead to either a reciprocal exchange of chromatid arms between homologs (crossover; CO) or a unidirectional replacement of a small genomic region in one of the chromatids, while leaving the donor homolog unchanged (noncrossover; NCO). (*b*) A schematic illustration of the CO detection in a three-generation pedigree. Highlighted by a star in the F_1 generation is the crossing over of grand-paternal (black) and grand-maternal (light gray) haplotypes. COs that occurred in the starred meiosis are detectable through sequencing all five individuals in the three-generation pedigree. (*c*) The four three-generation pedigrees (21 females and 11 males) that were used in this study. CO and NCO events were inferred in seven male F_1 individuals (color coded and highlighted by a star).

sex-averaged across male and female ancestors. Moreover, as LD-based approaches infer a population-scaled recombination rate ($\rho = 4 N_e r$, where N_e is the effective population size and r is the per-generation recombination rate), estimates can be confounded by population genomic processes that alter the levels of LD (such as the selective and demographic history of a population), if unaccounted for in the statistical model (Wall 2001 and see discussions in Dapper and Payseur 2018; Johri et al. 2022). To circumvent this issue, recently developed statistical methodologies explicitly account for the demographic history of the studied population when inferring recombination rates (e.g., Spence and Song 2019).

To gain further insights into the timescale, nature, and causes of recombination rate evolution, it is important to contextualize differences in the relative frequencies and distributions of CO and NCO events across taxa (Dapper and Payseur 2019). However, in contrast to COs, which have been a focal point of many studies in recent years (see the review of Stapley et al. 2017), detailed knowledge of NCO landscapes remains elusive in many species, despite their characterization being of vital importance to genome-wide association studies (Wall 2004), linkage analysis (Mancera et al. 2008), and positive selection scans (Yin et al. 2009), given their impact on local patterns of LD. Using a dense set of species-specific markers and explicitly accounting for population demography, we here utilize genome-wide single nucleotide polymorphism (SNP) data from four pedigrees of rhesus macaque (*Macaca mulatta*)—the most frequently used nonhuman primate model in biomedical research (Rogers 2022)—together with a recently released, highly contiguous rhesus reference genome assembly, in order to gain novel insights into the recombination landscape of the species. The application of a two-pronged approach—employing a direct pedigree-based method and an indirect demography-aware LD-based method—allowed us to obtain both a contemporary snapshot of genome-wide CO and, for the first time, NCO events as well as insights into the historical recombination rate of this animal model for human health and disease.

Materials and Methods

Data

Whole-genome data from 32 Indian-origin rhesus macaques (*M. mulatta*) spanning four three-generation pedigrees

housed at the California National Primate Research Center (21 females and 11 males; fig. 1c) previously sequenced to an average coverage of 39.5-fold per individual (supplementary table S1, Supplementary Material online) were downloaded from the NCBI Sequence Read Archive (BioProject PRJNA251548; Wang et al. 2020). Genetic relationships between individuals were confirmed using the software KING (Manichaikul et al. 2010) as implemented in plink2 (Chang et al. 2015). This demonstrated that the 11 individuals from the parental (*P*) generation were unrelated, except for one pair which displayed a relatedness of the third degree (supplementary fig. S1, Supplementary Material online).

Read Mapping

Following the Genome Analysis Toolkit (GATK) Best Practices (Van der Auwera and O'Connor 2020), adapter sequences were marked using GATK v.4.1.8.1 MarkIlluminaAdapters prior to mapping reads to the repeat-masked *M. mulatta* (rheMac10) reference assembly (GenBank accession number: GCA_003339765.3; Warren et al. 2020) using BWA-MEM v.0.7.17 (Li and Durbin 2009), flagging secondary alignments. After mapping, duplicates were marked using GATK v.4.1.8.1 MarkDuplicates, and multiple sequence realignments were performed around insertions/deletions (indels) using the GATK v.3.7.0 RealignerTargetCreator and IndelRealigner tools to improve alignments (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013). To detect, and correct for, systematic sequencing biases, base quality scores were recalibrated using the GATK v.4.1.8.1 BaseRecalibrator and ApplyBQSR tools, together with a training set of more than 70 million SNPs previously identified in 526 rhesus macaque individuals (Harris 2019), which led to more accurate base quality scores after recalibration (supplementary fig. S2, Supplementary Material online). Afterward, an additional round of duplicate marking was performed (as detailed above).

Variant Calling, Genotyping, and Filtering

Variants were called from high-quality mappings (mapping quality ≥40) using the GATK v.4.1.8.1 HaplotypeCaller, and individual call sets were combined using CombineGVCFs and jointly genotyped using GenotypeGVCFs. Thereby, the species-specific heterozygosity rate was set to 0.0024 (as previously reported by Warren et al. 2020), and the "--pcr-indel-model" was set to "NONE", as a PCR-free protocol was followed during sequencing (Wang et al. 2020).

The GATK Best Practices (Van der Auwera and O'Connor 2020) recommend to filter variants using the Variant Quality Score Recalibration (VQSR)—a sophisticated Gaussian mixture model that relies on the availability of an accurate SNP call set for training that is validated to a high degree of confidence. Despite having a large SNP dataset from a previous study on hand (Harris 2019), VQSR was not performed for three reasons. First, the publicly available SNP call set is based on an older reference assembly (rheMac8) that is of lesser quality than the most recent long-read assembly for the species (rheMac10). Regions that were less well-assembled in this older version could potentially lead to spurious variant calls. Second, variants were not experimentally validated. Third, although the dataset contained the chromosome, position, reference allele, and alternative allele for each SNP, annotations (INFO and FORMAT columns) that usually contain metrics related to the called genotypes (Danecek et al. 2011) had been removed from the Variant Call Format and replaced by a "." symbol, hindering an assessment of confidence in individual variant calls. As the accidental inclusion of erroneous SNP calls can lead to an inaccurate classification of variants during the VQSR step (see discussion in Li et al. 2019), variants were instead "hard" filtered following the GATK's Best Practice recommendations ("QualByDepth" < 2.0; "StrandOddsRatio" > 3.0; "FisherStrand" > 60.0; "RMSMappingQuality" < 40.0; "MappingQualityRankSumTest" < −12.5; "ReadPosRankSumTest" < −8.0; with acronyms defined by the GATK package; Van der Auwera and O'Connor 2020). To avoid potential biases resulting from computational imputation of genotypes, the dataset was further limited to autosomal, biallelic SNPs genotyped in all individuals. In addition, SNPs located within repetitive or low-complexity regions (as annotated by RepeatMasker v.4.0.8 [https://repeatmasker.org; last accessed July 2023] and Tandem Repeats Finder [Benson 1999] in the "soft-masked" reference assembly; Warren et al. 2020) were removed as these regions are difficult to reliably call with short-read sequencing data (see the review of Pfeifer 2017). Following Smeds et al. (2016), coverage and genotype quality filters were applied to minimize genotyping errors, retaining only those SNPs with a high confidence in the genotype (GQ ≥ 30; corresponding to a probability of an incorrect genotype of less than 0.001) that were covered by at least 15 reads but no more than twice the average autosomal coverage. In addition, SNPs exhibiting an excess of heterozygosity (determined using a *P*-value of 0.01 for Hardy–Weinberg equilibrium calculated using the "--hardy" option in VCFtools v.0.1.16; Danecek et al. 2011) as well as any sites violating Mendelian inheritance (determined using the GATK v.4.1.8.1 FindMendelianViolations tool) were removed.

The resulting call set contained 11,500,421 SNPs with a transition–transversion ratio of 2.2 (supplementary table S2, Supplementary Material online), distributed consistently across chromosomes (supplementary fig. S3, Supplementary Material online).

Power Analysis

To assess the power to identify SNPs carried by the 32 individuals included in this study, the final call set was

compared with a previously generated dataset containing more than 70 million SNPs identified in 526 rhesus macaque individuals (Harris 2019). Overall, 4.99% of SNPs were novel compared with the rhesus macaque panel, whereas 95.01% of SNPs were rediscovered (supplementary table S2, Supplementary Material online).

Pedigree-Based (Direct) Identification of CO and NCO Events

Three-generation pedigrees allow for the haplotype phasing of F_1 individuals and the identification of recombination events upon gamete transmission to the F_2 offspring (supplementary fig. S4, Supplementary Material online). Following Smeds et al. (2016), "phase-informative" sites were extracted from each three-generation pedigree using the following criteria. To trace an F_1 allele back to the parent from which it originated, the sire and dam must have distinct genotypes, and their F_1 offspring must be heterozygous. To reduce genotyping errors at heterozygous sites that could lead to spurious results, the alternative allele was required to be supported by at least 25% but no more than 75% of the reads. Moreover, to be able to trace the transmission of the F_1 allele to the F_2 offspring, either the F_1's partner or their joint offspring must be homozygous at the site. Once identified, the 5,573,111 phase-informative sites were then grouped into grand-maternal and grand-paternal haplotype blocks in the F_2 offspring using the pipeline illustrated in supplementary figure S4, Supplementary Material online.

COs were determined as genomic regions between the outmost SNPs of neighboring haplotype blocks in the F_2 offspring. Following Williams et al. (2015), individual phase-informative sites occurring within a region of SNPs with consistent phase were excluded to minimize genotyping errors. To reduce phasing errors and to account for CO interference (Otto and Payseur 2019), a minimum gap of 1 Mb was required between CO events within the same meiosis. Moreover, as, with the exception of humans (Miga et al. 2020; Nurk et al. 2022), highly repetitive telomeric regions remain poorly resolved in primate genome assemblies, CO events at the chromosome ends (within 2 Mb) were removed.

NCOs were determined as phase-informative sites that mismatched the surrounding haplotype block (i.e., two tightly linked CO events that happened within a single meiosis, or closely linked sites where an offspring inherited a haplotype block from one parent surrounded by a larger haplotype block from the other parent; Wall et al. 2022). To obtain a high-quality set of NCOs and to distinguish genuine NCOs from genotyping errors, phase-informative sites associated with either the presence of an insertion or deletion (within 10 bp) or a clustering of SNPs (defined here as three SNPs within a 10-bp region) as well as tracts with multiple haplotype changes within a 5-kb window were excluded.

Gene annotations obtained from the NCBI RefSeq collection (O'Leary et al. 2016) were used in ANNOVAR v.2014-07-14 (Wang et al. 2010) to categorize COs with a resolution of less than 5 kb and NCO events by genomic region (i.e., 5′ UTR, exonic, exonic noncoding RNA [ncRNA], intronic, intronic ncRNA, 3′ UTR, and downstream), and their distribution was plotted in R v.4.0.2 (http://www.R-project.org; last accessed July 2023).

LD-Based (Indirect) Estimation of Sex-Averaged Recombination Rates

Sex-averaged recombination rates were estimated from patterns of LD observed in the individuals from the parental (*P*) generation (supplementary table S1, Supplementary Material online) using the demography-aware recombination rate estimator pyrho (Spence and Song 2019). Following the pyrho guidelines, SMC++ v.1.15.5 (Terhorst et al. 2017) was first used to infer the size history of the population (using the "estimate" function) from segregating sites located on a single chromosome (chromosome 10), assuming a species-specific per-generation mutation rate of 5.8×10^{-9} per site (Wang et al. 2020). Model robustness was assessed by simulating five chromosomes of 99.5 Mb length (i.e., the length of chromosome 10) under the demographic model inferred by SMC++ with msprime v.1.2.0 (Kelleher et al. 2016) and comparing the expected folded site frequency spectra with the data (supplementary fig. S5, Supplementary Material online), excluding coding regions (defined as regions overlapping with, or present within 10 kb, of an exon) to mask genomic regions potentially experiencing purifying or background selection effects (following recommendations by Johri et al. 2021). Taking into account the demographic information inferred by SMC++, a likelihood lookup table was then generated in pyrho v.0.1.6 (using the "make_table" command), and the optimal parameter settings for window size and block penalty (smoothness) were determined (using "hyperparam"). Finally, using the recommended hyperparameter settings (i.e., a window size of 70 and a block penalty of 40), fine-scale recombination rates were estimated based on the patterns of LD observed in the data (using the "optimize" command; supplementary fig. S6, Supplementary Material online).

Results and Discussion

To study recombination rate landscapes in rhesus macaques (*M. mulatta*), high-coverage (average coverage = 39.5×; supplementary table S1, Supplementary Material online) whole-genome sequencing data of 32 Indian-origin individuals from four three-generation pedigrees (21 females and 11 males; fig. 1*c*) were used to detect 11.5 million autosomal SNPs (supplementary table S2, Supplementary Material online), 5.5 million of which were informative for

haplotype phasing. Tracing the transmission of alleles at phase-informative sites from the F_1 to the F_2 generations (fig. 1*b* and supplementary fig. S4, Supplementary Material online), a total of 171 autosomal COs were initially identified in the pedigrees. Studying the genomic distribution of these COs, three tightly clustered CO events originating from the same meiosis were detected within a 10-Mb window—a highly unlikely event, as CO interference usually prevents several COs in close proximity (Muller 1916). A similar pattern has recently been observed in a study of olive baboons (Wall et al. 2022). As we and others have discussed (Campbell et al. 2016; Wall et al. 2022), although such a pattern may indeed occur naturally, it is more likely driven by either genotype errors or a joint occurrence of a single CO event combined with an inversion (Broman et al. 1998, 2003), and thus, this region was excluded from further analysis. An additional region that showed indications of a misplaced contig (i.e., two or more COs from different meioses occurring in the same region) was also removed, resulting in a final dataset containing a total of 165 COs in seven paternal meioses (fig. 2), with a median resolution of 22.3 kb (supplementary table S3 and fig. S7*a*, Supplementary Material online). Consistent with a lower bound of one obligate CO per chromosome or chromosome arm necessary for the faithful segregation of homologs (Jones and Franklin 2006; Coop and Przeworski 2007), 0–3 COs per chromosome were observed within individual meioses (supplementary table S4, Supplementary Material online; note that COs only impact half of the gametes as illustrated in fig. 1*a*), leading to a total of 17–31 events per offspring (supplementary table S3, Supplementary Material online).

Based on the 165 CO events observed in seven meioses, the male-only autosomal genetic distance was estimated to be 2,357 cM (table 1), confirming earlier observations of a shorter genetic map length in rhesus macaques and other catarrhine monkeys (such as baboons and vervet monkeys) compared with anthropoid apes (humans, chimpanzees, bonobos, and gorillas; Rogers et al. 2000; Cox et al. 2006; Rogers et al. 2006; Jasinska et al. 2007; Auton et al. 2012; Pfeifer and Jensen 2016; Stevison et al. 2016; Pfeifer 2020; Xue et al. 2020; Wall et al. 2022; and see the review by Coop and Przeworski 2007). Specifically, an initial low-resolution linkage map for rhesus macaques utilized 241 human microsatellite loci genotyped in five pedigrees to directly estimate a sex-averaged autosomal genetic map of length 2,048 cM (Rogers et al. 2006); however, the authors suggested that this value is likely truncated by an estimated 500 cM due to a lack of molecular markers near the chromosome ends. Similarly, a cytological study that visually examined chiasmata (i.e., the physical representation of COs) in pachytene spermatocytes from rhesus macaques directly estimated the male-only autosomal genetic map length to be approximately 1,950 cM (Hassold et al. 2009), although the female recombination rate was hypothesized to be much higher. In agreement with the larger number of molecular markers (particularly at the chromosomal ends) and the higher quality reference assembly utilized in this study, the map presented here is 15% and 21% longer than these previous estimates (Rogers et al. 2006; Hassold et al. 2009).

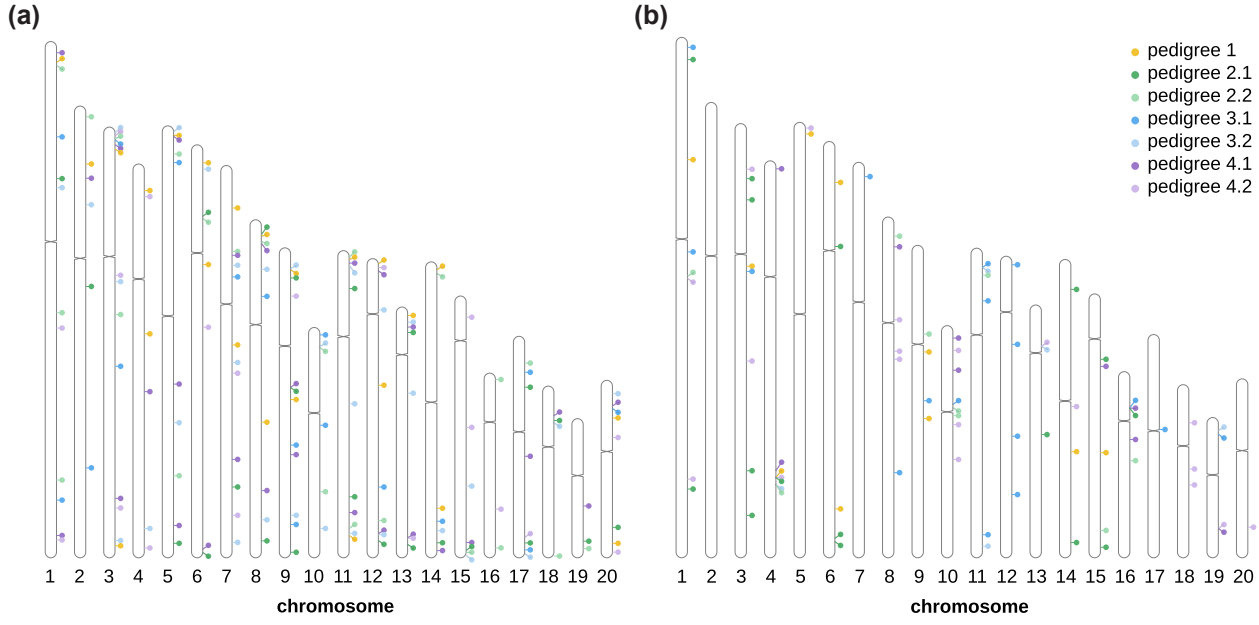

Fig. 2.—The genomic distribution of crossover and noncrossover events on autosomal chromosomes. Each circle represents (*a*) a crossover or (*b*) a noncrossover event from one of the seven paternal meioses used in the study (colored by pedigree following the coloring scheme introduced in fig. 1*c*).

Table 1
Autosomal recombination distance estimated from the number of observed CO events in the pedigrees

Chromosome	cM	cM/Mb
1	171.4	0.77
2	85.7	0.44
3	200.0	1.08
4	85.7	0.50
5	142.9	0.76
6	114.3	0.64
7	171.4	1.01
8	142.9	0.98
9	171.4	1.28
10	85.7	0.86
11	157.1	1.18
12	142.9	1.10
13	114.3	1.05
14	100.0	0.78
15	100.0	0.88
16	42.9	0.54
17	114.3	1.20
18	57.1	0.77
19	42.9	0.73
20	114.3	1.48
Autosomal	2,357.1	0.88

An estimate based on dense genome-wide SNP data from four pedigrees yielded a male-only autosomal genetic map length of 2,357 cM. Information regarding the relationship between the cumulative genetic distance obtained from the distribution of CO events and the physical length of each autosome is provided in supplementary figure S8, Supplementary Material online.

The average autosomal male CO rate in rhesus macaques was estimated to be 0.88 cM/Mb (table 1)—about 25% lower than the genome-wide average rate in humans (Hinch et al. 2011; Bhérer et al. 2017) but higher than the genome-wide average rates in Indian-origin rhesus macaques indirectly inferred utilizing patterns of LD observed in whole-genome population genetic data (ranging from 0.433 ± 0.333 cM/Mb in 100 kb windows to 0.448 ± 0.286 cM/Mb in 1 Mb windows; Xue et al. 2016, 2020). However, Xue et al. neglected to take the population history into account—an important caveat and potentially confounding factor, as the individuals included in their study originated from a severely bottlenecked research colony. Furthermore, as the conversion of their inferred population recombination rate into a per-generation rate required an assumption about the underlying effective population size—the estimation of which, in turn, relied on a human-like rather than a rhesus-like mutation rate (i.e., 10^{-8} rather than 5.8×10^{-9} per base pair per generation)—it is difficult to compare these earlier indirect estimates with that obtained here utilizing a more direct method. Moreover, these previous LD-based estimates (Xue et al. 2016, 2020) were based on the initial genome build for rhesus macaques (rheMac2) generated in 2006 (Rhesus Macaque Genome Sequencing and Analysis Consortium 2007), with subsequent research over more than a decade highlighting several sequencing errors and gaps in this draft assembly (e.g., Zhang et al. 2012; Norgren 2013) which likely resulted in a misinference of recombination rates in these regions.

Based on phase-informative sites that mismatched the surrounding haplotype block, 85 NCOs were identified in the four pedigrees (supplementary tables S3 and S5, Supplementary Material online)—fewer than the expected number of NCO events per generation in humans (95% confidence interval: 178–286; Williams et al. 2015). It should be noted though that, due to the stringent filtering necessary to prevent sequencing, genotyping, and phasing errors, this number likely represents a conservative estimate. Moreover, NCOs located between two markers as well as those repaired toward the original genotype will be missed in a pedigree-based analysis. As a consequence, the value presented here is likely an underestimate of the genome-wide frequency of NCO events in the species. Consistent with the tract lengths observed in other primates, including humans (55–290 bp; Jeffreys and May 2004; Odenthal-Hesse et al. 2014) and baboons (mean lengths: 42–167 bp; Wall et al. 2022), NCO tracts were generally short (mean length: 155 bp; supplementary table S3, Supplementary Material online), although two tracts longer than 1 kb were also detected (supplementary table S5 and fig. S7*b*, Supplementary Material online). The majority of NCOs (48 out of 85 or 56.5%) exhibited a pattern of GC-biased gene conversion, resulting in more GC-gametes than AT-gametes at AT/GC heterozygous sites (Galtier and Duret 2007; Duret and Galtier 2009)—a pattern also similar to that observed in humans (68%; Williams et al. 2015; Halldorsson et al. 2016) and baboons (57.6%; Wall et al. 2022).

In agreement with earlier work in humans and other mammals (Coop et al. 2008; Kong et al. 2010; Auton et al. 2012; Brick et al. 2012), COs and NCOs were reduced near transcription start sites, with a large proportion of events (50% and 47%) harbored in intergenic regions less frequently associated with promotor function (fig. 3). Moreover, out of the 20 COs overlapping genes, 95% were located in intronic regions. This bimodal distribution of CO events is consistent with the zinc finger protein PRDM9 (Baudat et al. 2010; Berg et al. 2010; Myers et al. 2010; Parvanov et al. 2010) being a major determinant of the location of meiotic recombination hotspots in non-human primates such as rhesus macaques (Oliver et al. 2009), which preferentially binds in intergenic regions as well as within actively transcribed genes (Walker et al. 2015).

To complement pedigree-based results, an LD-based sex-averaged recombination rate was inferred from the individuals of the parental generation (supplementary table S1, Supplementary Material online) using a demography-aware method to avoid confounding recombination rate estimates

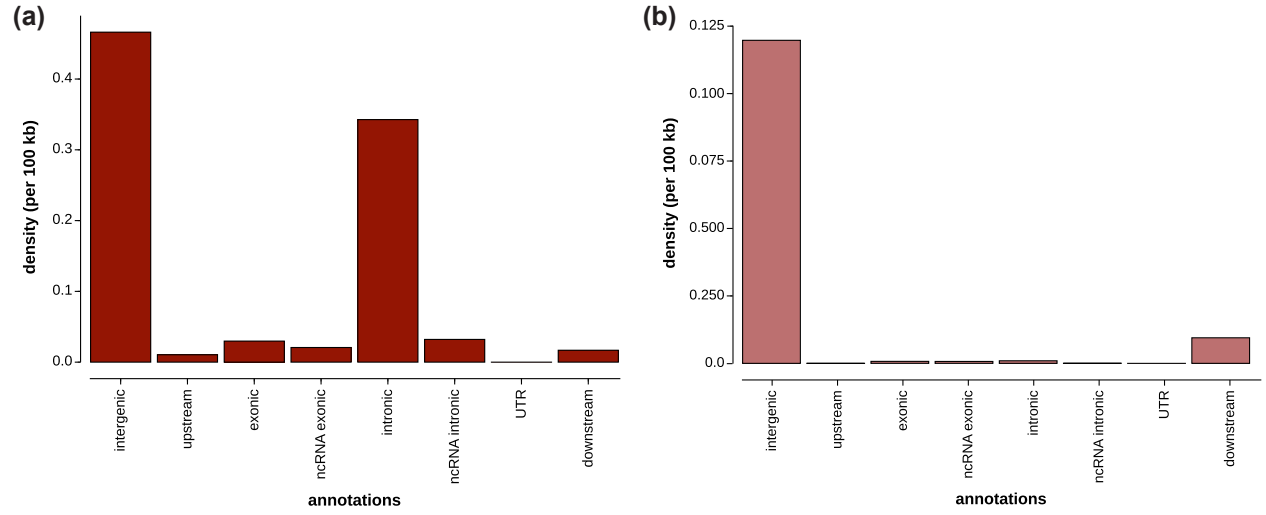

Fig. 3.—The density of crossover and noncrossover events by genomic region. The density of (*a*) crossovers with a resolution of less than 5 kb (shown in red) and (*b*) noncrossover events (salmon) by genomic region (i.e., 5′ UTR, exonic, exonic non-coding RNA [ncRNA], intronic, intronic ncRNA, 3′ UTR, and downstream). Genome annotations were obtained from the NCBI RefSeq collection (O'Leary et al. 2016).

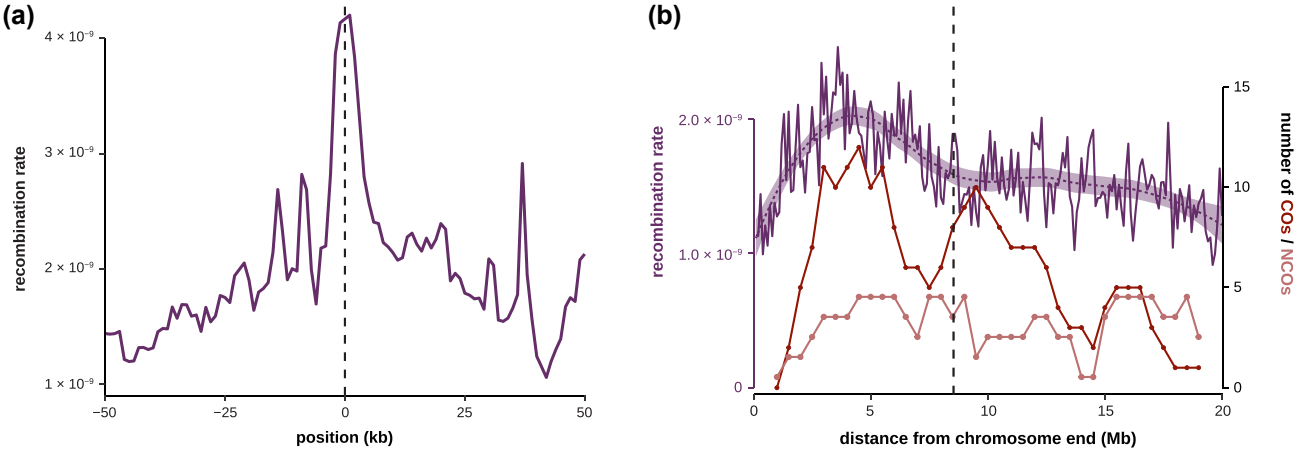

Fig. 4.—The distribution of recombination events. (*a*) Linkage disequilibrium-based recombination rate estimate at crossover locations (showing 50 kb regions upstream and downstream of the identified crossover site at position 0). (*b*) Number of paternal crossovers (COs; shown in red) and noncrossovers (NCOs; salmon) observed in the seven pedigrees as well as the historical sex-averaged recombination rate estimate indirectly inferred from patterns of linkage disequilibrium observed in unrelated individuals (purple) in relation to the distance from chromosome ends. CO and NCO counts were calculated in 2 Mb windows with a 250-kb step size, and recombination rates were estimated in nonoverlapping 100 kb windows (solid line), with a smoothed local regression (loess with span 0.5) shown as a dotted line.

with historical population size changes. LD-based recombination rate estimates were elevated at CO locations (fig. 4*a*), and both sex-averaged recombination rates and contemporary male CO, and to a lesser extent, NCO events showed a clustering near the chromosome ends (fig. 4*b*), consistent with other placental mammals (including other primates [Broman et al. 1998; Kong et al. 2002; Pratto et al. 2014; Venn et al. 2014; Wall et al. 2022], mice [Shifman et al. 2006; Cox et al. 2009; Brunschwig et al. 2012; Li et al. 2019], and dogs [Axelsson et al. 2012]).

Finally, it is important to acknowledge that the male-only CO and NCO landscapes that could be inferred from the pedigree data at hand provide a sex-biased view, as differences in recombination landscapes are known to exist between sexes in many mammalian species. Specifically, in many vertebrates, females exhibit more uniformly distributed CO landscapes with higher overall genome-wide rates than their male counterparts (see the review of Sardell and Kirkpatrick 2020). For instance, a recent pedigree-based study in another catarrhine monkey (olive baboons) found elevated CO rates near telomeres in males and reduced CO rates near telomeres in females, with an estimated overall male and female genetic map length of 2,080 and 2,506 cM, respectively (corresponding to a male-to-female

ratio of 1:1.2; Wall et al. 2022). In light of this, it is important to note that, due to the male-biased focus of this study, the reported 15–20% increase in genetic map length likely represents an underestimate of the sex-averaged rate. Consequently, to better integrate CO and NCO landscapes in this biomedically relevant species into genome-wide association studies or scans for positive selection, future work will need to focus on extending pedigree-based studies such as the one presented here not only to both sexes—which may also help reconcile the recombination rate observed in this catarrhine monkey with those observed in humans and other great apes—but also to individuals exhibiting different genetic backgrounds frequently utilized in research (e.g., Indian-origin vs. Chinese-origin rhesus macaques).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health ESI-MIRA award number R35GM151008 and the National Science Foundation CAREER award number DEB-2045343 to S.P.P. Computations were performed at the Arizona State University's High Performance Computing facility.

Data Availability

This study was based on whole-genome high-throughput sequencing data publicly available from the NCBI SRA (study accession number: PRJNA251548). The CO map generated in this study is included in the Supplementary Material online, and the LD-based recombination map is available at http://spfeiferlab.org/data.

Literature Cited

Auton A, et al. 2012. A fine-scale chimpanzee genetic map from population sequencing. Science 336(6078):193–198.
Axelsson E, Webster MT, Ratnakumar A, Ponting CP, Lindblad-Toh K. 2012. Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. Genome Res. 22(1):51–63.
Baudat F, et al. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science 327(5967):836–840.
Baudat F, de Massy B. 2007. Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. Chromosome Res. 15(5):565–577.
Baudat F, Imai Y, de Massy B. 2013. Meiotic recombination in mammals: localization and regulation. Nat Rev Genet. 14(11):794–806.
Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27(2):573–580.
Berg IL, et al. 2010. PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. Nat Genet. 42(10):859–863.
Bhérer C, Campbell CL, Auton A. 2017. Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales. Nat Commun. 8:14994.
Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. 2012. Genetic recombination is directed away from functional genomic elements in mice. Nature 485(7400):642–645.
Broman KW, et al. 2003. Common long human inversion polymorphism on chromosome 8p. In: Goldstein DR, editor. Science and statistics: a festschrift for Terry Speed. Vol 40. IMS lecture notes—monograph series. Beachwood (OH): Institute of Mathematical Statistics. p. 237–245.
Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. 1998. Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet. 63(3):861–869.
Brunschwig H, et al. 2012. Fine-scale maps of recombination rates and hotspots in the mouse genome. Genetics 191(3):757–764.
Buard J, de Massy B. 2007. Playing hide and seek with mammalian meiotic crossover hotspots. Trends Genet. 23(6):301–309.
Campbell CL, Bhérer C, Morrow BE, Boyko AR, Auton A. 2016. A pedigree-based map of recombination in the domestic dog genome. G3 (Bethesda). 6(11):3517–3524.
Chang CC, et al. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7.
Charlesworth B, Jensen JD. 2021. Effects of selection at linked sites on patterns of genetic variability. Annu Rev Ecol Evol Syst. 52:177–197.
Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134(4):1289–1303.
Clark AG, Wang X, Matise T. 2010. Contrasting methods of quantifying fine structure of human recombination. Annu Rev Genomics Hum Genet. 11:45–64.
Cole F, Keeney S, Jasin M. 2010. Comprehensive, fine-scale dissection of homologous recombination outcomes at a hot spot in mouse meiosis. Mol Cell. 39(5):700–710.
Cole F, Keeney S, Jasin M. 2012. Preaching about the converted: how meiotic gene conversion influences genomic diversity. Ann N Y Acad Sci. 1267:95–102.
Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in *Drosophila melanogaster*. PLoS Genet. 8(10):e1002905.
Coop G, Przeworski M. 2007. An evolutionary view of human recombination. Nat Rev Genet. 8(1):23–34.
Coop G, Wen X, Ober C, Pritchard JK, Przeworski M. 2008. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. Science 319(5868):1395–1398.
Cox A, et al. 2009. A new standard genetic map for the laboratory mouse. Genetics 182(4):1335–1344.
Cox LA, Mahaney MC, VandeBerg JL, Rogers J. 2006. A second-generation genetic linkage map of the baboon (*Papio hamadryas*) genome. Genomics 88(3):274–281.
Danecek P, et al. 2011. The variant call format and VCFtools. Bioinformatics 27(15):2156–2158.
Dapper AL, Payseur BA. 2018. Effects of demographic history on the detection of recombination hotspots from linkage disequilibrium. Mol Biol Evol. 35(2):335–353.
Dapper AL, Payseur BA. 2019. Molecular evolution of the meiotic recombination pathway in mammals. Evolution 73(12):2368–2389.
DePristo MA, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 43(5):491–498.
Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genomics Hum Genet. 10:285–311.

Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics 78(2):737–756.
Galtier N, Duret L. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. Trends Genet. 23(6):273–277.
Guryev V, et al. 2006. Haplotype block structure is conserved across mammals. PLoS Genet. 2(7):e121.
Halldorsson BV, et al. 2016. The rate of meiotic gene conversion varies by sex and age. Nat Genet. 48(11):1377–1384.
Harris AR. 2019. Rhesus macaque SNP calls VCF. Zenodo. Dataset. Available from: https://doi.org/10.5281/zenodo.3515522.
Hassold T, Hansen T, Hunt P, Vandevoort C. 2009. Cytological studies of recombination in rhesus males. Cytogenet Genome Res. 124(2): 132–138.
Hassold T, Hunt P. 2001. To err (meiotically) is human: the genesis of human aneuploidy. Nat Rev Genet. 2(4):280–291.
Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet Res. 8(3):269–294.
Hinch AG, et al. 2011. The landscape of recombination in African Americans. Nature 476(7359):170–175.
Jasinska AJ, et al. 2007. A genetic linkage map of the vervet monkey (*Chlorocebus aethiops sabaeus*). Mamm Genome. 18(5):347–360.
Jeffreys AJ, May CA. 2004. Intense and highly localized gene conversion activity in human meiotic crossover hot spots. Nat Genet. 36(2):151–156.
Johri P, et al. 2021. The impact of purifying and background selection on the inference of population history: problems and prospects. Mol Biol Evol. 38(7):2986–3003.
Johri P, et al. 2022. Recommendations for improving statistical inference in population genomics. PLoS Biol. 20(5):e3001669.
Jones GH, Franklin FC. 2006. Meiotic crossing-over: obligation and interference. Cell 126(2):246–248.
Kauppi L, Jeffreys AJ, Keeney S. 2004. Where the crossovers are: recombination distributions in mammals. Nat Rev Genet. 5(6): 413–424.
Kelleher J, Etheridge AM, McVean G. 2016. Efficient coalescent simulation and genealogical analysis for large sample sizes. PLoS Comput Biol. 12(5):e1004842.
Keeney S. 2001. Mechanism and control of meiotic recombination initiation. Curr Top Dev Biol. 52:1–53.
Kong A, et al. 2002. A high-resolution recombination map of the human genome. Nat Genet. 31(3):241–247.
Kong A, et al. 2010. Fine-scale recombination rate differences between sexes, populations and individuals. Nature 467(7319): 1099–1103.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760.
Li R, et al. 2019. A high-resolution map of non-crossover events reveals impacts of genetic diversity on mammalian meiotic recombination. Nat Commun. 10(1):3900.
Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature 454(7203):479–485.
Manichaikul A, et al. 2010. Robust relationship inference in genome-wide association studies. Bioinformatics 26(22):2867–2873.
Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet Res. 23(1):23–35.
McKenna A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20(9):1297–1303.
Miga KH, et al. 2020. Telomere-to-telomere assembly of a complete human X chromosome. Nature 585(7823):79–84.
Muller HJ. 1916. The mechanism of crossing over. Am Nat. 50: 193–221.
Muller HJ. 1964. The relation of recombination to mutational advance. Mutat Res. 1(1):2–9.
Myers S, et al. 2010. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. Science 327(5967):876–879.
Norgren RB. 2013. Improving genome assemblies and annotations for nonhuman primates. ILAR J. 54(2):144–153.
Nurk S, et al. 2022. The complete sequence of a human genome. Science 376(6588):44–53.
Odenthal-Hesse L, Berg IL, Veselis A, Jeffreys AJ, May CA. 2014. Transmission distortion affecting human noncrossover but not crossover recombination: a hidden source of meiotic drive. PLoS Genet. 10(2):e1004106.
O'Leary NA, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44(D1):D733–D745.
Oliver PL, et al. 2009. Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. PLoS Genet. 5(12): e1000753.
Otto SP, Payseur BA. 2019. Crossover interference: shedding light on the evolution of recombination. Annu Rev Genet. 53:19–44.
Parvanov ED, Petkov PM, Paigen K. 2010. *Prdm9* controls activation of mammalian recombination hotspots. Science 327(5967):835.
Pfeifer SP, Jensen JD. 2016. The impact of linked selection in chimpanzees: a comparative study. Genome Biol Evol. 8(10):3202–3208.
Pfeifer SP. 2017. From next-generation resequencing reads to a high-quality variant data set. Heredity (Edinb). 118(2):111–124.
Pfeifer SP. 2020. A fine-scale genetic map for vervet monkeys. Mol Biol Evol. 37(7):1855–1865.
Pratto F, et al. 2014. Recombination initiation maps of individual human genomes. Science 346(6211):1256442.
Przeworski M, Wall JD. 2001. Why is there so little intragenic linkage disequilibrium in humans? Genet Res. 77(2):143–151.
Rhesus Macaque Genome Sequencing and Analysis Consortium. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. Science 316(5822):222–234.
Rogers J. 2022. Genomic resources for rhesus macaques (*Macaca mulatta*). Mamm Genome. 33(1):91–99.
Rogers J, et al. 2000. A genetic linkage map of the baboon (*Papio hamadryas*) genome based on human microsatellite polymorphisms. Genomics 67(3):237–247.
Rogers J, et al. 2006. An initial genetic linkage map of the rhesus macaque (*Macaca mulatta*) genome using human microsatellite loci. Genomics 87(1):30–38.
Sardell JM, Kirkpatrick M. 2020. Sex differences in the recombination landscape. Am Nat. 195(2):361–379.
Shifman S, et al. 2006. A high-resolution single nucleotide polymorphism genetic map of the mouse genome. PLoS Biol. 4(12):e395.
Smeds L, Mugal CF, Qvarnström A, Ellegren H. 2016. High-resolution mapping of crossover and non-crossover recombination events by whole-genome re-sequencing of an avian pedigree. PLoS Genet. 12(5):e1006044.
Spence JP, Song YS. 2019. Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. Sci Adv. 5(10):eaaw9206.
Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM. 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philos Trans R Soc B Biol Sci. 372(1736):20160455.
Stevison LS, et al. 2016. The time scale of recombination rate evolution in great apes. Mol Biol Evol. 33(4):928–945.
Sun Y, et al. 2012. Deep genome-wide measurement of meiotic gene conversion using tetrad analysis in *Arabidopsis thaliana*. PLoS Genet. 8(10):e1002968.

Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat Genet. 49(2):303–309.
Van der Auwera GA, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 43(1110):11.10.1–11.10.33.
Van der Auwera GA, O'Connor BD. 2020. Genomics in the cloud: using Docker, GATK, and WDL in Terra. Sebastopol: O'Reilly Media.
Venn O, et al. 2014. Nonhuman genetics. Strong male bias drives germline mutation in chimpanzees. Science 344:1272–1275.
Walker M, et al. 2015. Affinity-seq detects genome-wide PRDM9 binding sites and reveals the impact of prior chromatin modifications on mammalian recombination hotspot usage. Epigenetics Chromatin. 8:31.
Wall JD. 2001. Insights from linked single nucleotide polymorphisms: what we can learn from linkage disequilibrium. Curr Opin Genet Dev. 11(6):647–651.
Wall JD. 2004. Close look at gene conversion hot spots. Nat Genet. 36(2):114–115.
Wall JD, Robinson JA, Cox LA. 2022. High-resolution estimates of crossover and noncrossover recombination from a captive baboon colony. Genome Biol Evol. 14(4):evac040.
Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38(16):e164.
Wang RJ, et al. 2020. Paternal age in rhesus macaques is positively associated with germline mutation accumulation but not with measures of offspring sociability. Genome Res. 30(6):826–834.
Warren WC, et al. 2020. Sequence diversity analyses of an improved rhesus macaque genome enhance its biomedical utility. Science 370(6523):eabc6617.
Williams AL, et al. 2015. Non-crossover gene conversions show strong GC bias and unexpected clustering in humans. Elife 4:e04637.
Xue C, et al. 2016. The population genomics of rhesus macaques (*Macaca mulatta*) based in whole-genome sequences. Genome Res. 26(12):1651–1662.
Xue C, et al. 2020. Reduced meiotic recombination in rhesus macaques and the origin of the human recombination landscape. PLoS One. 15(8):e0236285.
Yin J, Jordan MI, Song YS. 2009. Joint estimation of gene conversion rates and mean conversion tract lengths from population SNP data. Bioinformatics 25(12):i231–i239.
Zhang X, Goodsell J, Norgren RB. 2012. Limitations of the rhesus macaque draft genome assembly and annotation. BMC Genomics. 13:206.

Associate editor: Carolin Kosiol